

[Saadi Lahlou](#)

Modelisation et codage des aliments: theorie et application informatique pour l'observatoire des consommations alimentaires

**Article (Published version)
(Refereed)**

Original citation: Lahlou, Saadi (1993) *Modelisation et codage des aliments: theorie et application informatique pour l'observatoire des consommations alimentaires*. Bulletin d'Information de la Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes, 3 . pp. 94-115.

© 1993 Ministère de l'Économie, des Finances et de l'Industrie

This version available at: <http://eprints.lse.ac.uk/78618/>
Available in LSE Research Online: June 2017

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

MODELISATION ET CODAGE DES ALIMENTS :

THEORIE ET APPLICATION INFORMATIQUE POUR L'OBSERVATOIRE DES CONSOMMATIONS ALIMENTAIRES (DG:700)

Résumé :

La base de données de l'observatoire des consommations alimentaires (OCAbase) est un instrument qui doit fournir des courbes de répartition de la consommation de tous composants alimentaires : ingrédients, nutriments, contaminants, additifs, auxiliaires de fabrication.... Elle a été décrite en détail dans les publications techniques de l'observatoire (Lahlou et al.,/1991).

Pour dire vite, elle structure les données sous le schéma «des gens»/ «font quelque chose avec»/ «des choses à manger». Cette structure très générale permet d'inclure dans OCAbase des données d'enquêtes de consommation (des ménages/achètent/des produits ; des ménages/se procurent/des aliments), des enquêtes sur les prises alimentaires (des individus/ingèrent/des plats) ; etc.

Pour que cette base de données puisse absorber des données de provenance très diverse, il faut que le principe de codification de ces différents types d'objets (les gens, les actions, les choses à manger) soit très souple. Il faut aussi qu'il puisse être suffisamment fin pour qu'aucune information ne soit perdue lorsqu'on intègre ces données dans la base. Il a donc fallu mettre au point des modèles à portée très générale. On présente ici celui qui a été construit pour la partie «choses à manger» ;

Ce nouveau *modèle des aliments* est fondé sur une description des produits en termes de *facettes* et de *recette*. Ce système donne à OCAbase la même structure que le monde des aliments réels (chaque ingrédient, produit ou plat peut avoir son image dans la base). Cela facilite le passage entre les recherches dans la base de données et l'action en laboratoire ou sur le terrain.

En outre, pour permettre aux différents utilisateurs de continuer à utiliser leurs procédures habituelles, notre système de description intègre les systèmes existants, et en particulier Langua (Langua Alimentaria), le système de description des aliments mis au point par le Center for Food Safety and Applied Nutrition de

la Food and Drug Administration Nord-américain (CFSAN 1990 a et b), dont nous nous sommes beaucoup inspirés. Notre modèle de l'aliment se présente comme une sorte de sandwich, avec différentes couches de codage empilées.

Ce modèle puissant peut d'ailleurs s'appliquer à bien d'autres problèmes que celui de l'OCA.

Le désavantage est la lourdeur de l'opération de codification, qui risquerait de décourager les opérateurs d'avoir recours à notre système. Nous avons donc mis au point une *procédure de codification assistée par ordinateur*, qui s'inspire des mécanismes cognitifs spontanés de l'être humain et rend le codage à la fois rapide et fidèle. Son principe, que nous avons appelé codage par typicalité, consiste à coder chaque nouvel aliment en changeant à la marge la codification du produit déjà codé le plus proche («prototype»). Le prototype est repéré par le logiciel dans le réservoir de connaissance que constitue la base de données de l'observatoire ; il fait appel pour cela à l'analyse lexicale du nom courant (en langue naturelle) du produit à coder, qui est la seule information que le codeur doit fournir.

Ces principes ont été appliqués sous forme de logiciels. Le premier, CITOCA, réalisé en 1991, a permis de coder une dizaine de milliers de produits. Un second, CODIT, plus puissant et plus développé sur le plan de l'analyse linguistique, est en cours de perfectionnement.

Cet article décrit les principes de description et le processus de classification, il vise à montrer pourquoi nous pensons qu'il s'agit des solutions idoines. La description des logiciels proprement dite a déjà été faite pour CITOCA (Lahlou, Maffre, Beaudouin, 1992), et sera faite de façon plus détaillée dans les prochaines publications de l'OCA pour CODIT, qui ne sera ici présenté que de façon succincte dans sa version CODIT 1.0.

SOMMAIRE

1. COMMENT DECRIRE LES OBJETS ALIMENTAIRES

- 1.1. Chacun code midi à sa porte
- 1.2. Décrire, c'est réduire à peu
- 1.3. La description des objets
- 1.4. Façons de classer, façons de décrire
 - 1.4.1. *La description à l'aide de variables (qualités)*
 - 1.4.2. La description par nomenclature
- 1.5. Notre modèle de l'aliment

2. LA DESCRIPTION

- 2.1. Le choix des descripteurs
- 2.2. Les rubriques du codage
 - 2.2.1. *Libellé*
 - 2.2.2. Commentaires
 - 2.2.3. Langual
 - 2.2.4. INSEE, IFLS, Douanes
 - 2.2.5. Marketing
 - 2.2.6. Recette
 - 2.2.7. Codeur
 - 2.2.8. Origine
 - 2.2.9. Gestion interne

3. LA PROCEDURE DE CODIFICATION

- 3.1. L'objectif : traduire des descriptions naturelles dans le système des descriptions OCA
- 3.2. Le principe : codage par inférence et typicalité
- 3.3. L'application informatique

4. PROJETS

BIBLIOGRAPHIE

1. COMMENT DECRIRE LES OBJETS ALIMENTAIRES ?

1.1. Chacun code midi à sa porte

Connaître la consommation alimentaire des gens est essentiel pour de nombreux opérateurs : nutritionnistes, toxicologues, économistes, juristes, industriels... à des fins de recherche, de réglementation, de santé publique, de marketing. Chacun de ces opérateurs a de l'aliment une perception particulière qui découle de ses besoins. Ainsi, certaines des facettes de l'objet alimentaire qui sont pertinentes pour l'homme de marketing (la marque) seront inintéressantes pour le nutritionniste, qui, lui, s'intéressera à d'autres aspects (composition...).

Chaque opérateur, lorsqu'il recueille des données sur les aliments, se limite, pour des raisons évidentes, aux seuls aspects qui l'intéressent. Chaque profession, chaque secteur, se constitue des systèmes descriptifs adaptés à ses besoins. Ainsi, en France cohabitent actuellement divers systèmes de codification des aliments : la nomenclature alimentaire INSEE (Institut National de la Statistique et des Etudes Economiques), utilisée pour l'enquête alimentaire nationale ; la nomenclature activité produits (NAP) qui sert notamment en comptabilité nationale, la nomenclature des Douanes, la nomenclature de l'institut français du libre service (IFLS), le code à barres... qui sont des systèmes de codage nationaux.

Plus localement, les nutritionnistes, quand ils cherchent à utiliser une description normalisée, se réfèrent aux catégories de produits utilisées dans les tables de composition qu'ils utilisent, mais même ces tables ne sont pas standardisées. Les différentes sociétés d'étude de marché ont des systèmes de codification différents entre elles, voire d'une enquête à l'autre dans un même institut. Dans les autres pays européens la situation est analogue.

Pratiquement rien n'est fait pour harmoniser ces différentes nomenclatures ; et c'est compréhensible, puisqu'une nomenclature unique ne saurait présenter qu'un point de vue et serait donc inutilisable pour les autres opérateurs. A l'extrême, si l'on voulait rassembler tous les points de vue, on arriverait à des classifications comme celle que Borges (1957, p. 144) fait attribuer par le docteur Franz Kuhn à «certaine encyclopédie chinoise» intitulée «le marché céleste des connaissances bénévoles» où

les animaux se divisent en a) appartenant à l'Empereur, b) embaumés, c) apprivoisés, d) cochons de lait, e) sirènes, f) fabuleux, g) chiens en liberté, h) inclus dans la présente classification, i) qui s'agitent comme des fous,

j) innombrables, k) dessinés avec un très fin pinceau de poils de chameau, l) et contra, m) qui viennent de casser la cruche, n) qui de loin semblent des mouches».

Cette situation dans laquelle chacun définit son objet suivant ses préoccupations propres n'est, en soi, pas gênante. Elle n'empêche, après tout, que de se communiquer des données.

Or, justement, ce partage est devenu nécessaire. D'abord, parce que les données sont coûteuses à recueillir : il est souhaitable qu'elles puissent servir plusieurs fois, et donc que leur codage puisse servir à plusieurs, et si possible à tous. Ensuite, il est de plus en plus nécessaire de faire des comparaisons entre enquêtes, au niveau national ou européen : pour des raisons de coût, les grandes études doivent souvent s'appuyer sur plusieurs enquêtes locales, un peu comme une couverture en patchwork.

C'est en particulier le problème qui s'est posé pour l'observatoire. OCABase est constituée à partir des données tirées de diverses enquêtes sur la consommation alimentaire, notamment celle de l'INSEE, du panel SECODIP (société d'études de la consommation, distribution, publicité), et bientôt des données de consommations individuelles tirées d'enquêtes nutritionnelles. Non seulement les données proviennent d'opérateurs ayant chacun des objectifs différents quand ils collectent les informations, mais aussi les utilisateurs sont multiples et ont des intérêts divers. Ces intérêts concernent la consommation, la santé publique, la réglementation, la nutrition, l'orientation de la production... Il nous faut donc disposer d'une description des produits alimentaires suivant des dimensions qui soient compatibles avec ces champs d'intérêts. C'est difficile car la codification est par essence réductrice.

1.2. Décrire, c'est réduire à peu

La codification détruit l'objet, ou, plus exactement, elle le réduit aux seules dimensions du système de codification. Prenons un exemple fictif. Brillat-Savarin écrit dans sa physiologie du goût (p. 152) :

«J'ai parcouru les dictionnaires au mot gourmandise, et je n'ai point été satisfait de ce que j'y ai trouvé. Ce n'est qu'une confusion perpétuelle de la gourmandise proprement dite avec la gloutonnerie et la voracité : d'où j'ai conclu que les lexicographes, quoique très estimables par ailleurs, ne sont pas de ces savants aimables qui embouchent avec grâce une aile de perdrix au suprême pour l'arroser, le petit doigt en l'air, d'un verre de vin de Laffite ou du Clos-Vougeot».

Le codeur qui voudrait décrire ce que mangent ces aimables lexicographes avec la nomenclature INSEE traduirait sans doute :

*aile de perdrix au suprême : INSEE53 (gibier)
verre de vin de Laffite ou de Clos-Vougeot : IN-
SEE9113 (vin d'appellation d'origine contrôlée)*

Celui qui voudrait retrouver, à partir de ces codes, ce qu'a consommé le gastronome en question ne pourra plus jamais revenir à l'objet initial, il n'a que les codes INSEE53 et INSEE9113 qui pourraient tout aussi bien signifier

rôti de sanglier et Riesling

Car la perte d'information est irrémédiable : on ne peut pas entièrement reconstruire un objet à partir de sa seule description. Un nom d'objet n'est pas un objet. On ne peut jamais, à partir de la seule carte, savoir exactement ce qu'est le réfectoire qu'elle décrit. Ainsi, chacun sait que «plateau de fromages» ne renvoie pas à la même chose sur la carte de Bocuse et sur la carte du restaurant d'à côté. Derrière le code «produit pour pâtisserie» (INSEE863) y avait-il de la levure, ou du sucre vanillé ?

Il faudrait que chaque opérateur décrivant un produit puisse en coder une description très détaillée, afin que l'objet une fois codé soit appauvri le moins possible. Mais il ne faut pas que cette description extensive soit trop longue ou trop compliquée à coder, sinon personne ne l'utilisera. Alors, comment décrire ?

1.3. La description des objets

Le monde est infiniment divers, et chaque objet est unique. On pourrait écrire sur chaque objet une monographie fort longue, bornée seulement par les capacités de description de l'observateur.

Or, agir dans un monde où tous les objets sont considérés comme irréductiblement différents en tous points exclut la possibilité de transférer la connaissance que l'on a d'un objet à un autre. Notre système scientifique tire son efficacité de sa capacité à opérer des généralisations, des regroupements d'objets en catégories, c'est-à-dire en ensembles ayant au moins une propriété en commun. On peut alors résumer en lois applicables à des catégories les résultats de nos expériences. Pour nous, penser, c'est donc d'abord classer ; et comprendre, c'est d'abord reconnaître.

Le classement consiste à assigner à un objet une position dans le système de description du monde de l'observateur. L'observateur qui classe cherche à transformer des descriptions des objets telles qu'elles lui sont initialement fournies (par ses propres sens, par des descriptions indirectes) en coordonnées dans son propre système de références.

Il y a une infinité de façons de décrire tout objet, il y a une infinité de systèmes de références possibles. Parmi toutes ces façons, certaines sont plus pertinentes dans un contexte donné. Ces façons pertinentes se caractérisent par un usage immédiat pour l'action qui nous intéresse : ne perdons jamais de vue que la raison pour laquelle nous créons des catégories est toujours utilitaire. Les descriptions, et les classements (qui sont des descriptions particulières) sont donc des façons de voir les objets qui sont adaptées à un certain contexte d'utilisation. Elles sélectionnent, parmi les descriptions possibles de l'objet, un certain nombre de dimensions pertinentes, ou facettes, dans chacune desquelles un objet est repéré par une caractéristique, ou descripteur.

Le problème est délicat parce qu'il n'existe pas, dans l'absolu, de description définitive des objets.

«(...) les sciences exactes de la nature partent de l'idée que, finalement, il sera toujours possible de comprendre la nature, dans chaque domaine nouveau de l'expérience ; mais comme on ne l'a pas fixé a priori le sens du terme «comprendre», la connaissance de la nature, formulée mathématiquement par des époques antérieures, bien que «définitive», n'est pas toujours applicable. (...) Le concept de vérité scientifique, fondement des sciences, peut comporter de nombreuses façons de comprendre la nature.» (Heisenberg, p. 32).

Par exemple, si les facettes sont : la famille, la marque, le poids, on pourra décrire un objet comme :

chocolat-noir-en-tablettes/Lindt/100-grammes

chacun des termes séparés par des «/» étant un descripteur correspondant à une facette.

Si les dimensions sont : le goût, la couleur, la consistance, on pourra décrire un objet comme :

chocolaté/brun-foncé/fondant

Si les dimensions sont : ingrédient principal, matériau de contact d'emballage, état physique dans les conditions standard de température et de pression, on peut décrire un objet tel que :

cacao/papier-d'aluminium/solide

Il n'existe pas a priori de solution générale optimale pour les systèmes de description mais seulement des solutions «idoines», qui ne peuvent être déterminées qu'empiriquement. Pour expliquer pourquoi, il nous faut évoquer ici une question difficile, celle de la nature de l'objet décrit. Elle ne sera pas résolue, et nous n'y reviendrons plus par la suite ; mais elle éclaire ce qu'est la limite épistémologique du processus de codification.

A propos des trois descriptions du chocolat (supra), rien, évidemment, ne nous permet d'affirmer a priori que ces différentes descriptions sont différents points de vue d'un même objet.

Malgré sa trivialité apparente, cette affirmation soulève des problèmes ontologiques extrêmement délicats ⁽¹⁾, beaucoup débattus depuis longtemps. Ceux-ci portent en dernier ressort sur la validité du postulat de réalité (« existe-t-il une nature immanente, dont la forme s'impose à l'entendement humain, ou est-ce l'observateur qui crée le monde à l'image de ses structures cognitives ? »). La raison utilitariste (comprendre pour agir) de l'acte de décrire renvoie la description de l'objet à un contexte qui les détermine. La définition d'un objet est une opération qui implique beaucoup plus que l'objet lui-même (à supposer qu'un tel objet puisse être isolé de son contexte).

Une des façons d'ouvrir la porte sur l'abîme ontologique est de chercher à préciser l'origine des données : s'agit-il d'échantillon à doser (objet unique et concret), d'un poste de nomenclature dans une enquête consommateurs (catégorie formelle), d'une déclaration par un ménage (énoncé en langue naturelle)... ? Peut-on manger le poste « chocolat en tablettes » de l'enquête INSEE ? Peut-on additionner deux tablettes, ou bien additionne-t-on les quantités mesurées des occurrences attribuées à un même type ?

Ces questions naïves ne sont pas innocentes, car, si chacun sait qu'il est interdit d'ajouter des pois et des carottes, objets de nature différente, il nous faudra cependant bien, dans notre cas concret de l'observation, additionner des kilos de pois et de carottes pour parler de kilos de légumes, surtout quand ils ne nous apparaissent que comme composants indifférenciés de l'objet décrit comme :

jardinière-de-légumes/surgelés/Picard/500g

Et il va falloir que nous sachions de quoi on parle pour fournir des statistiques sur les « légumes ».

Car nous cherchons, finalement, à répondre à des questions dans un certain univers (par exemple, celui des nutritionnistes), à partir de données obtenues dans un autre (celui des panélistes). Ce que le nutritionniste voit comme assemblage de nutriments, le consommateur le voit comme « nourriture », le statisticien le voit comme « poste de nomenclature », et le panéliste le recueille comme « Unité de Vente au Consommateur ». Aucune catégorie n'est exactement

traduisible en une autre, car les représentations et les usages qui les sous-tendent sont de nature différente. Là où le consommateur voit « un steak », le nutritionniste distingue cent qualités de viande différentes. Là où le statisticien enregistre « un kilo de pommes de terre », le consommateur réalise cent recettes. Là où le nutritionniste voit « un camembert à 45 % de matière grasse », le panéliste repère cinquante marques différentes, et le consommateur quatre gradations de mûrissement. Là où le législateur voit « une bouteille de vin de Bordeaux AOC », le distributeur voit des milliers de références différentes, le prix pouvant varier de un à dix ou plus selon l'année pour un même cru. Les limites de ces différentes catégories ne coïncident pas.

Certaines catégories pertinentes pour un opérateur n'existent même pas pour certains autres. Par exemple, les appellations « garnitures », « grossissants » (Benguigui, 1973), « produits légers de petit déjeuner » (Lahlou, 1988), (catégories consommateur) ; ou « sucres lents » (concept nutritionniste), « vins doux naturels » ou « ultra-frais » (catégories distributeur), n'évoqueront rien à certaines classes d'acteurs. Le consommateur distinguera mal la limite entre les fromages frais et certains fromages blancs, pour lesquels la réglementation cependant diffère. L'économiste, l'homme de marketing, et d'une façon générale chaque acteur de la filière perçoit des caractéristiques différentes (Lahlou, 1985).

Il est, par construction, impossible de fournir une solution exacte à une question impliquant un choix de catégorisation (par exemple : « combien un consommateur français consomme-t-il de lait ? »). Toute réponse sera une approximation, parce qu'il faudrait s'entendre sur ce qu'est « lait ». Faut-il compter la poudre de lait utilisée en biscuiterie ? Et avec quel type d'équivalence de mesure quand il s'agit de poudre de lait écrémé ? Pour préciser en quoi et combien c'est une approximation, il nous faut impérativement expliciter les opérations de traduction que nous faisons d'un univers dans l'autre. Comme l'écrit Heisenberg à propos de la connaissance de la matière au niveau atomique :

« Pour les sciences de la nature également, le sujet de recherche n'est donc plus la nature en soi, mais la nature livrée à l'interrogation humaine et dans cette mesure l'homme, nouveau, ne rencontre ici que lui-même » (Heisenberg, p. 29).

(1) Par exemple : à supposer qu'il existe un tel objet, serait-ce l'échantillon analysé ou la catégorie ? Qu'est-ce qui nous permet de distinguer le type de l'occurrence de type dans le cadre d'une description empirique où le type est le résultat d'une induction ? (Par exemple : « L'analyse a révélé la présence de *Listeria* »). Le lecteur intéressé par les questions d'identification des variables qualitatives (accord sur les noms de phénomènes) pourra se reporter par exemple à l'argumentation de W.V.O. Quine (1977) sur « lapin/Gavagai », à Wittgenstein (1961, 1965). Pour les rapports entre mesure et dimensions sur les variables quantitatives, voir par exemple la première partie de G. Monod-Hertzen (1976) et H. Saget (1981).

Nous ne pouvons donc pas nous contenter de la commode position scientifique «classique», qui consiste à négliger la relativité des données. Cependant, pour ne pas aller trop loin dans le relativisme, nous proposons de camper sur une position acceptable par tous, qui consiste à poser la relativité de toute mesure, et de toute description des phénomènes, sans chercher à trancher sur la qualité ontologique des phénomènes eux-mêmes.

Poirier (1981) donne un aperçu très clair de la manière dont on peut garder une position opérationnelle en faisant à la position idéaliste les concessions indispensables. Par exemple :

*«(...) presque toutes les mesures sont indirectes et se font dans le cadre et par l'intermédiaire de toute une théorie, ce qui ne veut aucunement dire qu'elles soient conventionnelles mais que les réponses se font dans le langage conceptuel des questions, et que le système des questions et des réponses s'adapte solidai-
rement à l'expérience.*

(...) à l'inverse du monde des objets sensibles, qui a une sorte d'unicité absolue, le monde des objets théoriques a une pluralité d'aspects, suivant les systèmes conceptuels à l'aide desquels nous le décrivons ou le contruisons. C'est ainsi que nous n'avons qu'une manière de voir un objet, mais diverses manières de le peindre ou de le dessiner, suivant les procédés, les techniques, les styles que nous mettons en œuvre.

(...) la réalité objective, c'est ce qu'il est nécessaire ou simplement plausible d'admettre pour sauver les phénomènes, dans un langage cohérent.

Ce qu'il faut admettre, c'est que plusieurs langages sont possibles pour décrire un objet en lui-même mystérieux, sans qu'il soit possible de dire quel est le vrai, puisqu'aucune expérience ne peut les départager, et que par ailleurs ils correspondent en général à des orientations plutôt qu'à des contenus définis de la pensée».

Nous ne chercherons donc pas à résoudre la question de la nature des produits alimentaires. Nous cherchons simplement à mettre au point un système cohérent de traduction, de codification et de classement des descriptions des objets alimentaires. Nous ne cherchons pas à décrire les produits pour notre propre usage, mais à permettre à chacun de comprendre et d'utiliser pour lui-même les descriptions d'objets faites par tous les autres.

1.4. Façons de classer, façons de décrire

Décrire c'est traduire le langage concret du monde en langage humain. Du tissu confus du phénomène, combinaison inintelligible de signes, nous extrayons un ensemble de traits, qui, nommés et combinés, en constituent une description compréhensible.

«(...) ce monde fut créé par combinaison de possibles qui, dans l'entendement divin, forment éternellement les notes en ségrégation (sejuncta), lettres ou nombres, d'une table qui convient à la nôtre en certains rapports. (...).

Or, nous trouvons tous ces éléments «pêle-mêle», multitude de «détails» infiniment complexe, qui répond originairement à ce que le phénomène mondial a de «multiplex». Science des complexions élémentaires, la combinatoire est, pour Dieu, science des possibles et organon de la constitution du monde, elle est, pour nous, doctrine de déchiffrement de l'univers (...).» (Serres, pp. 105-107, passim).

La description d'un objet dans une langue ne peut se faire que par l'énonciation structurée d'une suite de mots. Toute description est donc fatalement une combinaison de mots. Or, les mots de la langue, d'une manière ou d'une autre, renvoient à d'autres objets du monde supposés déjà bien connus de l'observateur, ou à des catégories abstraites. Une description compréhensible se ramènera finalement à une combinaison d'objets bien connus par l'observateur.

Il existe chez l'homme deux grands types d'approches cognitives des objets. La première approche est celle en termes de prototypes, et l'autre en termes de qualités. Pour dire vite, l'approche en termes de prototypes consiste à décrire un objet en le comparant à un autre, on lui attribue alors un type unique auquel il est assimilé. L'approche en terme de qualités est une approche plus analytique : elle consiste à décrire un objet par des propriétés. La description par types suppose un modèle relativement indépendant de l'observation. La description par qualités est plus liée au processus d'observation : les qualités sont des dimensions sensibles de l'appareil d'observation.

Ces deux approches, souvent concomitantes, supposent une organisation préalable du monde de référence, soit en catégories d'objets-repères (prototypes), soit en catégories de description (qualités). Il existe forcément une «encyclopédie du monde», dont le codeur suppose l'existence chez le destinataire de la description, et qui lui permettra de comprendre son message.

La puissance d'un système descriptif réside dans la capacité à pouvoir marquer et distinguer facilement un grand nombre d'objets différents avec un petit nombre de traits et un petit nombre de règles simples. Il n'existe pas de système idéal, tout dépend de ce qu'on cherche à décrire : si l'on distingue beaucoup de propriétés mais peu de catégories, mieux vaut décrire par catégories : sinon, c'est l'inverse. Le premier cas est adapté à des objets chacun très complexe, mais très typés. Par exemple, dans une ville, on aura à distinguer des hommes, des voitures, des bâtiments. Dans le cas où les objets varient continûment selon quelques critères seulement, il vaudra mieux se fonder sur ces critères. Par exemple, on utilisera pour décrire une population de ballons la taille, la couleur, la matière.

1.4.1. La description à l'aide de variables (qualités)

Une variable est un symbole (X , t , M , etc...) qui peut prendre toutes les valeurs d'un ensemble appelé le domaine de la variable. Si ce domaine est réduit à une valeur, la variable est dite constante. L'idée sous-jacente est de décrire l'objet en extension sous chacune des dimensions qui paraît, a priori, observable et pertinente à l'observateur. Ce système a l'avantage de la souplesse. Les variables sont un système efficace pour comparer les objets.

Par contre, ce système n'est pas économique en termes de classement, car chaque objet doit être décrit par un grand nombre de variables pour commencer à avoir une forme compréhensible. Par exemple :

couleur : jaune
diamètre : 5 cm
forme : sphérique
usage : comestible
origine : biologique

décrit probablement un fruit ou un légume, mais il faudrait beaucoup plus de variables pour nous amener à en avoir une idée précise.

1.4.2. La description par nomenclature

Pour avoir une définition courte de l'objet, et pour se communiquer des résultats, on utilise souvent le système de la «nomenclature», dans lequel il existe a priori un nombre fini de catégories, dans lesquelles on peut classer tous les objets, d'une façon univoque. Les nomenclatures hiérarchiques autorisent plusieurs niveaux de classement imbriqués plus ou moins fins. L'intérêt de la nomenclature est que chaque poste se réfère à un objet ayant une certaine forme générale, que l'observateur est supposé connaître.

Par exemple, le terme «prune», à lui seul, nous en dit beaucoup plus sur l'objet que la description «jaune,

sphérique, 5 cm, comestible, d'origine biologique». Ceci, à condition bien sûr que nous connaissions le prototype en question ; «rambutan» sera sans doute moins évocateur pour les observateurs français. C'est la première limite de ce système. La seconde est qu'il ne permet pas de comparer les objets de types différents, mais seulement de compter des objets de chaque type.

Poussé à l'extrême, ce système de description cherche à englober tous les objets dans une unique variable hiérarchisée. C'est ce qui s'est passé dans les sciences de la nature, comme en botanique à partir des travaux de Linné, ou dans le commerce avec la nomenclature des douanes. C'est également ce qui se fait dans la description usuelle des produits alimentaires, notamment dans les enquêtes de l'INSEE.

Il est possible de renoncer totalement à un tel système qui, si l'on en croit Pécerc, «ne marche pas» à un niveau global. Mais comme nous avons cependant besoin de classer pour penser, il nous faudrait alors construire des classifications *ad hoc* pour chaque problème spécifique à partir d'une description simplement en termes de variables non hiérarchiques. C'est possible en utilisant des techniques mathématiques de classification ou de segmentation ; Cela ne facilite pas la communication, puisqu'alors chaque classification locale est idiomatique. Nous semblons, en théorie, condamnés à perdre d'un côté ce que nous gagnons d'un autre.

Ce constat, pour vrai qu'il soit, n'est pas constructif, et nous avons besoin d'un outil, fut-il imparfait. Alors, dans la pratique, confrontés à un vaste ensemble d'objets, sur lesquels on veut réaliser des travaux statistiques très variés, comment peut-on construire un système de description pratique, puissant et flexible ? Comment réaliser concrètement ce système, et établir des correspondances avec d'autres systèmes descriptifs ?

1.5. Notre modèle de l'aliment

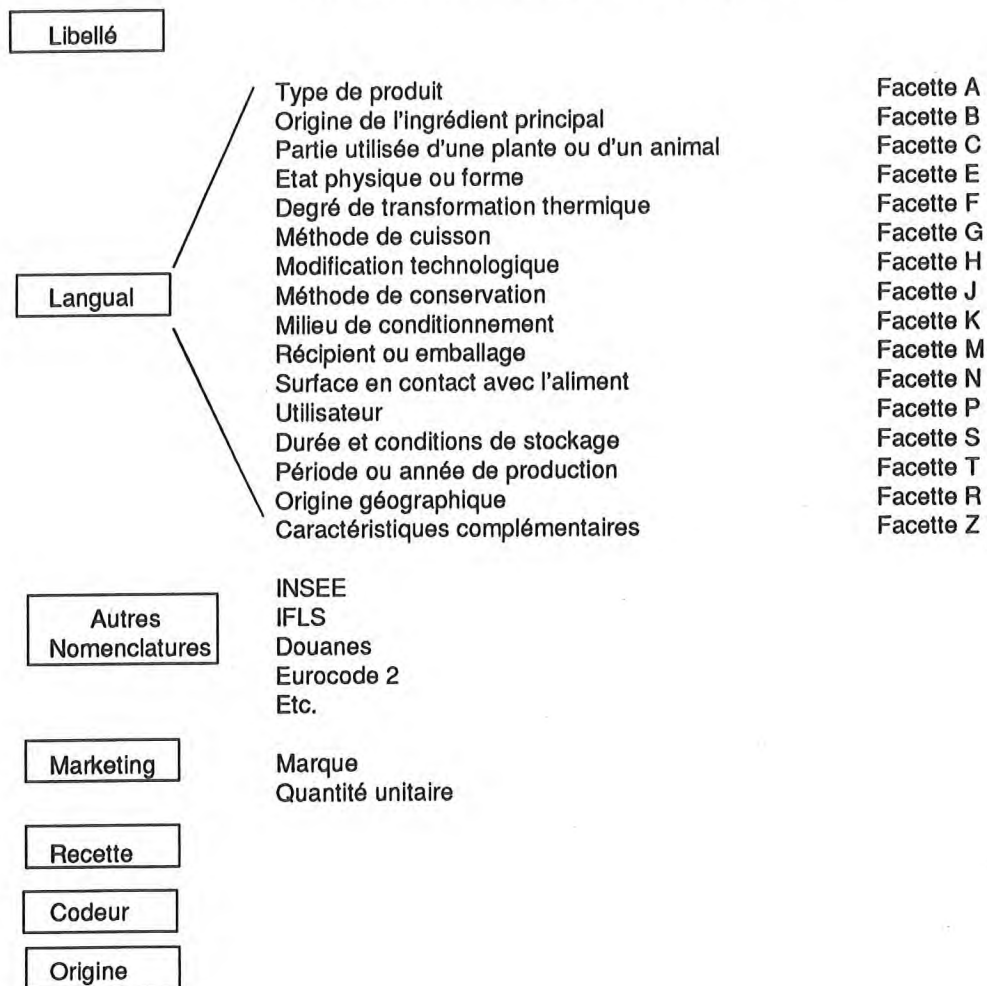
Notre idée de base est simple. Nous allons, d'une part, décrire l'objet dans chacun des langages spécialisés des utilisateurs potentiels. Il s'agit là d'une simple *concaténation* des vues existantes, une sorte de dictionnaire multilingue qui donne de l'objet un nom dans le langage de chacun des utilisateurs. Cette description par concaténation des codes existants consiste à décrire l'objet *en tant que tel* par une approche à travers une série de nomenclatures hiérarchiques. Chaque nomenclature est considérée comme une variable. On combine ainsi les avantages analytiques de la description par variables, et les avantages pragmatiques de la description classificatoire. On a donc ici une approche classique, plus compliquée que les descriptions usuelles, puisque nous avons juxtaposé plusieurs systèmes descriptifs, mais essentiellement identique.

Mais nous allons également décrire les aliments dans leur langage propre, où chaque aliment sera décrit comme une combinaison d'autres aliments. Nous avons construit le modèle de description en considérant chaque objet alimentaire comme une entrée de dictionnaire. Chaque aliment est alors, comme dans une encyclopédie, un objet en tant que tel, mais il a également une expression définitoire dans le langage des objets alimentaires. Dans la pratique, nous avons adopté le principe très simple des «recettes». Les recettes sont connues depuis toujours comme moyen opératoire d'agir sur les aliments, en les combinant par une grammaire culinaire.

On introduit ici une structure des objets, indépendante du processus d'observation, qui retrace le processus

de construction expérimentale (fabrication) de ces objets. Notre système descriptif y gagne là une caractéristique essentielle, celle d'avoir une structure homomorphe à celle du monde réel ; il reproduit dans sa structure de description les opérations de constitution des objets du monde réel : les aliments sont décrits à l'image de la façon dont ils sont fabriqués, par assemblage. Cette idée ambitieuse tranche avec tous les principes antérieurs de codification alimentaire ; elle n'est rendue applicable que par la puissance considérable des systèmes modernes de traitement de l'information par les bases de données relationnelles. Ceci est réalisé dans la pratique grâce à un choix apparemment anodin, celui de placer les ingrédients et les recettes sur le même plan formel : tout ingrédient est une recette, et toute recette peut être un ingrédient. ⁽²⁾

Le modèle de l'aliment «sandwich à facettes»



(2) On applique ici les principes développés dans la «formalisation en relativité complète» basée sur l'analyse combinatoire (Lahlou, 1990). L'idée de l'analogie entre recette alimentaire et représentation du monde n'est pas nouvelle : elle a été largement développée en anthropologie par les structuralistes, et d'abord Levi-Strauss ; chez les logiciens, on remarquera que, dans sa «méréologie» (formulation logico-mathématique d'une ontologie fondée sur l'opération d'inclusion), Lesniewski avait utilisé la notion «d'ingrédient» comme notion de base.

Entendons-nous : Il faut bien distinguer les *composants des ingrédients*. Les ingrédients sont des objets alimentaires qui, par l'opération d'une *recette*, ont formé le produit. Par exemple, le beurre, la farine, les œufs et le sucre sont les ingrédients du gâteau. Les composants sont le résultat de l'analyse, ex-post, de l'aliment lui-même. Le gâteau sera ainsi décomposé en protéides, lipides, glucides, résidus de pesticides, colorants naturels, etc. Les composants sont des *résultats d'analyse*, issus d'un travail de *description*, et non pas de construction de l'aliment. L'énoncé d'une composition, par exemple dans une table nutritionnelle, a le même statut qu'un codage dans une nomenclature : c'est le résultat d'une opération scientifique de l'observateur sur l'objet alimentaire. Chaque composant est décrit par la mesure d'une variable continue. La description d'une recette a un autre statut ; c'est la combinaison d'ingrédients par un fabricant, pour créer

l'aliment. Chaque ingrédient est un objet à part entière, qui a le même statut ontologique que l'objet codé en tant que recette. C'est bien sûr la *combinaison des ingrédients dans une recette* qu'est fondé notre système de description ; et non pas sur l'analyse d'une composition.

La puissance de notre modèle provient de ce que les utilisateurs ont une visée pragmatique (se rapportant au monde réel). Quel que soit leur intérêt sur l'aliment (toxicologie, santé publique, production...), une fois que la connaissance est acquise, celle-ci doit se transformer en action, et pour cela il faut pouvoir repérer les aliments dans leur processus de transformation, depuis la matière première agricole jusqu'au plat ingéré. Il fallait donc que ce processus transparaîsse dans la description des aliments.

Un exemple : la crème Chantilly en bombe

NOM (NAME)		CREME CHANTILLY (BOMBE)	CHANTILLY CREAM
LANGUAL	A0269 B1201 C0154 E0130 F0001	NAPPAGE OU FOURRAGE VACHE CREME A FOUETTER LIQUIDE NIVEAU INCONNU DE TRANSFORMATION THERMIQUE	SWEET SAUCE OR TOPPING COW WHIPPING CREAM LIQUID EXTENT OF HEAT TREATMENT NOT KNOWN
	G0003	PAS DE METHODE DE CUISSON APPLICABLE	COOKING METHOD NOT APPLICABLE
	H0136	SUCRE OU SIROP DE SUCRE AJOUTE	SUGAR OR SUGAR SYRUP ADDED
	H0178 J0108	AERE TRAITE PAR DES CONSERVATEURS CHIMIQUES	AERATED PRESERVED BY TREATMENT WITH CHIMICALS
	K0015	CONDITIONNE EN BOMBE AEROSOL	PACKED WITH AEROSOL PROPELLANT
	M0168 N0042 P0024	TUBE EN ALUMINIUM ALUMINIUM ALIMENT POUR HUMAINS SANS PRECISION D'AGE	ALUMINIUM TUBE ALUMINIUM HUMAN FOOD NO AGE SPECIFICATION
	R00FR	FRANCE	FRANCE
	INSEE	7511	CREME UHT ET CREME CHANTILLY
	IFLS	9221	CREME FRAICHE
	MARQUE Brand name	XXXX	XXXX
RECETTE Recipe41	12305	CREME FRAICHE SUCRE	CREAM SUGAR
POIDS NET Net weight		125 g	125 g

Sur le plan théorique, notre modèle présente trois avantages :

- il est pragmatique jusque dans sa structure formelle ;
- il est compatible avec les façons de voir le monde actuelles des opérateurs ;
- il est analogue à la langue naturelle et par conséquent facile à maîtriser par les humains.

Sur le plan pratique, ce modèle présente deux inconvénients :

- la description de chaque objet est très longue, et donc lourde à coder ;
- la description en termes d'ingrédients n'est possible qu'avec l'encyclopédie des définitions réciproques de tous les ingrédients.

Cependant, notre modèle de l'aliment peut être traduit avec des outils mathématiques bien connus et programmables : théorie des ensembles, analyse combinatoire, théorie des graphes. Il est parfaitement adaptable au fonctionnement des bases de données relationnelles.

2. LA DESCRIPTION

2.1. Le choix des descripteurs

Le système de description nous permet de décrire tous les observables, il ne comporte pas en lui-même de procédé qui permette de savoir si les objets codés existent ou non. Tout ce qu'il peut vérifier est la bonne forme des descriptions, sur le plan syntaxique.

La structure générale de description est un vecteur :

(libellé, commentaire, Langua, INSEE, IFLS, douanes, marketing, recette, codeur, origine, gestion interne).

Un objet sera décrit comme un vecteur particulier, dans lequel chaque variable a pris une modalité particulière. La syntaxe n'est donc pas libre, et le vocabulaire non plus.

On a choisi de segmenter le codage en sections étanches, c'est-à-dire que chacune des rubriques du code (libellé, Langua, etc.) peut être codée indépendamment des autres, et dispose de sa propre syntaxe et de son propre vocabulaire :

2.2. Les rubriques du codage

2.2.1. Libellé

C'est le nom du produit en langue naturelle, correcte ou dégradée. Par exemple :

baba au rhum, recette maison

ou

*Beghin-Say sucre blanc en morceaux n° 3
boîte carton 1 kg*

La longueur de ce nom est cependant limitée à 170 caractères.

2.2.2. Commentaire

On peut ici insérer, en texte libre, des précisions éventuelles sur l'aliment. Cette facette sert à caser ce qui n'aurait pas pu être rentré par ailleurs dans le codage. Par exemple :

orthographe du nom incertaine

2.2.3. Langua

Langua (Langua Alimentaria) est la description de l'objet en termes physiques mis au point par le CFSAN (Center for Food Safety and Applied Nutrition, organisme de recherche dépendant de la Food and Drug Administration des Etats-Unis d'Amérique). Langua est une codification combinatoire comportant plus de 3 000 descripteurs élémentaires, sorte de thesaurus de traits qui servent à décrire les objets.

Pour faciliter la codification, l'archivage et la recherche, ces *descripteurs* sont eux-mêmes classés en grandes catégories, dimensions ou « facettes », qui correspondent à de grandes classes de propriétés. Par exemple, des *descripteurs* comme « solide », « liquide », « semi-liquide », « émulsion », « gaz » seront rassemblés dans une *facette* appelée « état physique ». On trouvera dans la bibliographie des références d'articles sur ce système, utilisé en France par le Cical (centre informatique sur la qualité des aliments), et qui est actuellement le système de description des aliments le plus abouti et le plus performant (Feinberg, Irelan-Ripert et Favier, 1991).

Le système Langua est un ensemble de descripteurs composés d'un code et d'un libellé organisés selon seize catégories ou facettes :

Type de produit	Facette A
Origine de l'ingrédient principal	Facette B
Partie utilisée d'une plante ou d'un animal	Facette C
Etat physique ou forme	Facette E
Degré de transformation thermique	Facette F
Méthode de cuisson	Facette G
Modification technologique	Facette H
Méthode de conservation	Facette J
Milieu de conditionnement	Facette K

Récipient ou emballage	Facette M
Surface en contact avec l'aliment	Facette N
Utilisateur	Facette P
Durée et conditions de stockage	Facette S
Période ou année de production	Facette T
Origine géographique	Facette R
Caractéristiques complémentaires	Facette Z

tème permet l'ajout ou la modification de codes sans que la structure globale en soit affectée.

Les facettes correspondent à des classes de traits définitoires.

La définition en Languag de la tomate se présente comme suit :

Chaque facette comprend un nombre variable de codes qui correspondent à toutes les «valeurs» que peut prendre la facette. A titre d'exemple voici un extrait du contenu possible de la facette E (état physique ou forme) :

- E0130 - liquide
 - E0102 - liquide très-visqueux
 - E0139 - liquide très-visqueux sans particule visible
 - E0121 - liquide très-visqueux avec de petites particules
 - E0138 - liquide très-visqueux avec des morceaux solides
 - E0109 - liquide peu-visqueux
 - E0123 - liquide peu-visqueux sans particule visible
 - E0114 - liquide peu-visqueux avec de petites particules
 - E0149 - liquide peu-visqueux avec des morceaux solides
- E0001 - état physique ou forme inconnu
- E0108 - état physique ou forme multiple
- E0103 - semi-liquide
 - E0110 - semi-liquide avec des morceaux solides
 - E0135 - semi-liquide à consistance lisse
- E0144 - semi-solide
 - E0134 - semi-solide avec des morceaux solides
 - E0119 - semi-solide à consistance lisse

L'information est résumée par un code composé d'un identifiant de la facette et d'une suite alphanumérique. C'est un principe partiellement hiérarchique qui organise chaque facette. La structure ouverte de ce sys-

11529	Tomate fraîche
A0152	légume
B1276	Tomate (<i>Lycopersicum esculentum</i>)
C0139	fruit non pelé avec trognon ou noyau ou graine
E0150	entier de forme naturelle
F0003	sans transformation thermique
G0003	pas de méthode de cuisson applicable
H0003	aucun traitement appliqué
J0001	traitement de conservation inconnu
K0003	sans milieu de conditionnement
M0001	récipient ou emballage non spécifié
N0001	surface en contact avec l'aliment inconnue
P0024	produit de consommation courante
R0	
S0	
T0	
Z0	

Le nom est décrit par une combinaison de traits. Mais alors qu'une définition en langage naturel répond à la seule règle de l'économie (seuls les traits pertinents apparaissent), la description en Languag, faite pour être utilisée par des machines aussi, comporte toujours seize rubriques - certaines pouvant éventuellement ne pas être remplies. La structure de la description est orientée vers des usages spécifiques propres à certains univers (nutritionniste, médical, technologique...). La souplesse du système permet d'apporter une information plus ou moins exhaustive selon les cas et d'avoir la possibilité de remplir progressivement les facettes.

Chaque facette contient un nombre variable de descripteurs, qui peuvent éventuellement être hiérarchisés (jusqu'à 10 niveaux dans la facette B).

**Nombre de descripteurs par facettes
dans LANGUAL, fin 1991**

FACETTES	NOMBRE DE POSTES
A Type de produit	196
B Origine de l'ingrédient principal	1435
C Partie utilisée d'une plante ou d'un animal	163
E Etat physique ou forme	46
F Degré de transformation thermique	7
G Méthode de cuisson	33
H Modification technologique	264
J Méthode de conservation	49
K Milieu de conditionnement	39
M Récipient ou emballage	112
N Surface en contact avec l'aliment	47
P Utilisateur	33
R Origine géographique	371
S Durée et conditions de stockage	3
T Période ou année de production	2
Z Caractéristiques complémentaires	136
TOTAL	2936

Notons que la facette H (modification technologique) tient plus ou moins lieu de recette dans Languel, puisqu'elle peut contenir un nombre non limité de descripteurs comme «sucre ajouté» ou «épice ou herbe ajouté». Nous conservons cette facette afin de rester totalement compatible avec Languel, bien que notre système de codification des recettes donne également ces informations.

2.2.4. INSEE, IFLS, Douanes...

Pour les parties du sandwich qui sont des nomenclatures, nous utilisons chacune des nomenclatures existantes exactement telle qu'elle est, en conservant ses descripteurs et sa structure, mais de façon totalement étanche aux autres nomenclatures. Il y a simplement juxtaposition de ces nomenclatures. Par exemple, pour la nomenclature INSEE, nous utiliserons exactement les codes INSEE. Ces termes garderont entre eux leurs relations sémantiques originelles (notamment les différents niveaux de nomenclature), à l'intérieur de la tranche de code correspondante, mais n'auront aucune relation avec les autres couches de code.

2.2.5. Marketing

Cette rubrique contient des informations réparties en 3 facettes : la marque, la quantité unitaire, et le «par combien» (lorsque les produits sont vendus en lots).

Nous créons les descripteurs de ces facettes rubrique au fur et à mesure que de nouveaux cas se rencontrent.

2.2.6. Recette

La recette, telle qu'elle est codée, dans le produit, est un simple numéro d'ordre qui renvoie à une table des recettes.

Une recette est un tableau constitué d'une superposition de vecteurs, dont chacun correspond à un ingrédient de la recette. Ce vecteur est constitué du numéro de l'ingrédient, de la quantité utilisée, de l'unité de cette quantité (poids, pourcentage...), et du traitement technologique appliqué à cet ingrédient dans la recette.

N° du 1er ingrédient	quantité	Unité	Traitement technologique
N° du 2ème ingrédient	quantité	Unité	Traitement technologique
N° du 3ème ingrédient	quantité	Unité	Traitement technologique
N° du 4ème ingrédient	quantité	Unité	Traitement technologique
...			

Le N° d'ingrédient renvoie à un produit existant dans la base et déjà codé en tant que produit. Le traitement technologique est une opération simple (cuisson, battage...). C'est une variable que nous sommes en train de construire en nous inspirant de Languel, et dont l'élaboration n'est pas terminée. Les codages en cours sont à blanc pour cette variable.

Notons qu'un ingrédient simple est une recette dont il est lui-même l'unique ingrédient.

2.2.7. Codeur

Cette facette repère l'instance qui a réalisé le codage (personne, organisme). Elle permet de remonter à la source de l'information si nécessaire.

2.2.8. Origine

Cette facette désigne l'origine de l'objet codé. Il peut s'agir par exemple d'un produit déclaré dans une enquête auprès des ménages, ou provenant de la déclaration d'un fabricant, ou d'une recette provenant d'un livre de cuisine, d'un échantillon ayant fait l'objet d'une analyse...

2.2.9. Gestion interne

Cette rubrique comprend divers codes qui permettent de gérer la base de données, de vérifier son intégrité, d'assurer des procédures de sécurité. Par exemple, y figurent le numéro du prototype à partir duquel le produit a été codé, sa date de codification etc...

3 LA PROCEDURE DE CODIFICATION

3.1 L'objectif : traduire des descriptions naturelles dans le système des descriptions OCA.

Nous allons nous intéresser à des objets qui sont eux-mêmes des descriptions d'objets. C'est un problème de traduction. Notre matériau de base est constitué de données recueillies, ou construites, par d'autres : nous n'avons qu'une connaissance de deuxième main des phénomènes. Par exemple, une référence SECODIP sera :

20401102030500/indus filets poisson/colin lieu noir/de 300 à 499 g

et une référence INSEE :

6121 poissons surgelés

Quand nous travaillons sur des libellés (noms de produits), nous ne possédons que cette description ; quand nous travaillons directement sur le produit, en général son fabricant ou son distributeur lui ont déjà donné un nom, et des caractéristiques qu'il nous faut prendre en considération.

Dans une certaine vision du monde, les objets sont perçus comme ayant un certain nombre de traits, qui font partie du système de description propre à cette vision du monde. En langue naturelle, ces traits sont signifiés par des mots (noms-de-traits), que l'on appelle encore, dans certains jargons, «libellés» ou «codes», et que nous désignons du nom général de descripteurs.

Par exemple, dans une vision du monde mercatique, le trait dont une traduction possible serait

«a été fabriqué par la société SODAP-Nestlé»

sera signifié par le libellé

«fabricant : Nestlé»

et, dans un contexte non ambigu, abrégé en

«Nestlé»

«Nestlé» est alors le libellé signifiant le trait. Si, dans un répertoire, les fabricants ont été repérés, non par des caractères alphabétiques, mais par des nombres, et que le code correspondant au trait

«a été fabriqué par la société SODAP-Nestlé»

est

05489

alors, «05489» sera une autre façon de symboliser le trait en question et sera donc une traduction du libellé «Nestlé» dans un autre langage. On peut encore dire qu'il lui correspondra.

Les objets sont définis par des combinaisons de traits. La description de l'objet en tant que tel se fait par les combinaisons des descripteurs correspondantes. La combinatoire des traits définit une classe d'objets observables, obtenus par la combinaison des traits élémentaires, que l'on appellera univers de cet observateur.

Dans cet univers, seules certaines combinaisons existent en fait, c'est ce que nous appellerons le monde de cet observateur. Par exemple, l'univers de SECODIP contient des objets tels que :

saumon-fumé/morceaux-N°3/pack-de-12/Kellogs/allégé/vanille.

qui est bien une combinaison de traits SECODIP. Mais un tel objet ne fait pas partie du monde des observables de SECODIP. C'est une description possible, mais qui ne «correspond à rien», de même que

une lampe à refroidir en odeur de licorne

est une description en bon français, mais qui ne correspond pas à un objet attesté par une observation empirique.

Un objet comme :

A0152/B1276/C0229/E0121/F0014/G0003/
H0151/J0123/K0003/M0001/N0001/P0024

autrement dit

légume-ou-dérivé/tomate-(Lycopersicum-esculentum)/fruit-pelé-sans trognon-ni-noyau-ni-graine/liquide-très-visqueux-avec-de-petites-particules/transformation-thermique-complète/pas-de-méthode-de-cuisson-applicable/épicé/stérilisé-à-chaud-ou-mis-en-conservation/sans-milieu-de-conditionnement/récipient-ou-emballage-non-spécifié/surface-en-contact-avec-l'aliment-inconnue/produit-de-consommation-courante

existe dans le monde de REGAL, nomenclature du centre informatique pour la qualité des aliments (CI-QUAL) où il porte le n° 11549, et de Languel, mais pas dans l'univers de SECODIP. Dans le monde courant, une définition plus compréhensible nous est donnée par son libellé dans REGAL : «sauce tomate en conserve».

La codification va consister à traduire une description naturelle de l'objet, provenant du monde de l'observateur qui veut le coder, vers le système de description que nous venons de décrire (le sandwich à facettes). Le système que nous présentons ici permet de traduire des descriptions initiales en langue naturelle correcte ou dégradée. Il n'est pas conçu pour faire de la transcription directe de code à code.

Nous cherchons à trouver des règles de traduction, qui nous permettront, si nous trouvons dans SECODIP un objet comme :

00100011010215/THE CEYLAN CARTON
AVEC SACHETS LIPTON

de lui attribuer un certain nombre de caractéristiques pour nous intéressantes dans les autres univers, comme par exemple de savoir que ce produit est obtenu à partir du *Thea sinensis*, et ainsi le transformer en un objet comme :

A0268-infusion/B1623-thé (*Camellia* ou *Thea sinensis*)/C₂₀₀-feuille/E0150-entier/de forme naturelle/F0003-sans transf thermique/G0003-pas de méth de cuisson applicable/H0138-déshydraté/J0116-déshydraté ou séché/K0003-sans milieu de conditionnement/M0197-sac ou sachet papier/N0039-carton ou papier/P0024-produit de consommation courante/LK-SRI LANKA/0/0

3.2 Le principe : codage par inférence et typicalité

Pour coder, nous allons avoir recours à deux méthodes. La première est celle de la codification directe, ou inférentielle, qui consiste à attribuer directement des codes à l'objet à partir de la compréhension de sa description (son libellé) en utilisant des règles d'inférence.

En l'occurrence, on pourra directement ici reconnaître la marque («LIPTON», dans l'exemple précédent). Il va pour cela falloir utiliser des informations sur les propriétés des objets, sous la forme de règles qui, à partir de la lecture du libellé, permettent de déduire directement certaines caractéristiques du produit alimentaire. On peut par exemple reconnaître facilement la marque, la quantité unitaire, le type de conditionnement, ou encore certaines grandes catégories (plats cuisinés, boissons...). Dans un monde donné, pour opérer une classification, on peut procéder en appliquant des règles de classement qui ont pour prémisses des traits observés. Par exemple :

Canard de Barbarie surgelé Findus

sera codé en

volaille surgelée

car «tout canard est une volaille», et que «tout surgelé est un surgelé». Ces opérations permettent de remplir automatiquement une partie du «club sandwich» de codage. Leur principe a été détaillé dans Lahlou, Maffre et Beaudouin (1992).

La seconde méthode de codage consiste à reconnaître dans l'objet à coder un objet proche d'un autre objet déjà codé, et à lui attribuer les codes de cet objet prototype, en les modifiant éventuellement à la marge. C'est la codification indirecte ou typicaliste. C'est un procédé de classement «indirect», car il utilise non plus les propriétés de l'objet lui-même, mais ses relations avec d'autres objets déjà classés. On pourra alors classer B «de la même façon que l'on a classé A», sous certaines conditions reliant B et A. Il suffit de disposer de relations entre les observables, par exemple une distance, une mesure de similarité. Par exemple (fictivement) on classera

Canard de Barbarie Vivagel

en

volaille surgelée

parce que

«*Canard de Barbarie Vivagel*»

est du même type que

Canard de Barbarie surgelé Findus

qui lui même est classé en

volaille surgelée.

Cette méthode, quoique peu immédiate sur le plan théorique, est très couramment appliquée dans la pratique. C'est la méthode naturelle de classification des objets nouveaux. Ainsi, Linné, lors de son voyage en Laponie, en 1732, à Vasterbotn :

«Le 7 juin. Je parlais de Gränon très tôt le matin (...). Je vis ici un oiseau étrange, qui jusqu'à maintenant n'est pas décrit. Si je m'en souviens bien, le prof. Rudbeck l'appelait Pica Laponum³; je l'examinais avec la longue-vue et lui trouvais toutes les notas d'une grive, adeoque non vereor eum Turdus cauda rubra, medio cinerea⁴. En plus, il poussa un cri et cela ressemblait à celui d'un Turdus, il volait comme un Turdus, etc. vide descriptionem quam in Ornithologia conscripsi.»⁵ (Linné, p. 51).

Le succès de cette méthode tient, pensons-nous, à ce qu'il est relativement plus facile de fabriquer des «comparateurs», fonctions ou processus dont les arguments portent sur une même classe d'objets, que des «identificateurs», qui sont capables de décrire des traits sur un objet. A tel point que les «identificateurs» rencontrés empiriquement en biologie sont en général des «comparateurs» à un modèle donné : l'identification est alors (comme on pouvait étymologiquement le prévoir) une re-connaissance. Il est peut-être même possible que l'identification par comparaison soit, après tout, la méthode canonique, si, comme le prétend Bateson l'information est l'expression d'une différence. Et, de fait, l'examen détaillé des procédures concrètes de ce que nous appelons «classement direct» risque de découvrir un enchaînement de classements indirects.

Quoi qu'il en soit, ce mécanisme efficace doit être envisagé dans un système de classification qui se veut opératoire. Nous pouvons alors fonder sur ce mécanisme ce que nous appellerons la codification typicaliste, qui consiste à nous référer à un prototype pour coder un objet donné. Cette approche a par ailleurs l'avantage de correspondre à une stratégie cognitive naturelle (voir par exemple les travaux d'Eleanor Rosch).

Le prototype de référence n'existe que dans notre vision du monde, c'est en fait un modèle. Mais il peut éventuellement être considéré comme parfaitement réalisé en une occurrence particulière. Ainsi, on pourra considéré que

«sucre blanc en morceaux n° 3»

est parfaitement réalisé dans

*«Beghin-Say sucre blanc en morceaux n° 3
boîte carton 1 kg».*

Si nous disposons d'une description de ce prototype dans le langage qui nous intéresse, nous pouvons alors partir de celle-ci pour décrire d'autres objets qui sont proches de ce prototype, en modifiant à la marge les descripteurs qui ne conviennent pas, un peu à l'image des criminologues qui cherchent à reconstituer des visages à partir de bandes photos des yeux, nez, menton, etc. Le fait de savoir «en gros» à quoi ressemble l'objet que l'on cherche à décrire permet de gagner un temps considérable.

3.3. L'application informatique

Deux logiciels⁶ successifs ont été construits sur les principes décrits ci-dessus.

Le premier, CITOCA, a été réalisé en 1991 et a permis de coder une dizaine de milliers de produits alimentaires issus du panel SECODIP. Ce logiciel a été développé avec le logiciel de gestion de bases de données 4ème Dimension sur Apple Macintosh. (Lahlou, Mafre et Beaudouin, 1992). Très convivial en raison de son interface graphique, ce logiciel restait cependant limité par les possibilités de son langage de programmation et par la puissance des machines sur lesquels il pouvait être installé.

C'est pourquoi nous avons commencé à développer un logiciel plus puissant, qui inclut des fonctionnalités de reconnaissance linguistique, et présente l'avantage d'être directement relié à OCABase. Ce logiciel, CODIT, prend également en compte les recettes de façon beaucoup plus performante, et a des liens avec le système d'interrogation assisté que nous avons développé pour la base (THESAURUS). Ces deux logiciels ont été développés avec le langage de 4ème génération ABF et le générateur de grilles d'écran VIFRED de INGRES. Dans sa version actuelle, CODIT 1.0 a les mêmes fonctionnalités que CITOCA, mais son architecture est prévue pour un développement plus important (multilinguisme, contrôles de cohérence du codage, codification en série...)⁷. THESAURUS et CODIT 1.0 sont interfacés à OCABase et correspondent chacun à une centaine de milliers de lignes de programmes en langage C. CODIT a déjà

(3) *Pica Lapponum* : Pie de Laponie

(4) *adeoque... cinerea* : et pour cela je n'hésite pas à la nommer grive à queue rouge, le milieu cendré

(5) *vide... conscripsi* : voir la description faite par l'ornithologie

(6) A parler rigoureusement, il s'agit d'applications informatiques et non pas de logiciels.

(7) CODIT a été développé notamment par Michael Krus, Gloria Calamassi-tran et Ghislaine Cathenod THESAURUS par Gérard Verheyden, Gloria Calamassi-Tran, Ghislaine Cathenod, Christopher Russel et Michael Krus.

permis de coder environ 15 000 références alimentaires du marché français à partir des libellés du panel SECODIP, dans une version restreinte du «sandwich à facettes» de codage OCA.

L'utilisation de CODIT 1.0. est simple. Le codeur tape en clair le libellé du produit qu'il souhaite coder.

ex : Yaourt lait entier nature pot en verre

Le logiciel recherche dans OCABase, parmi les produits déjà codés, ceux dont le libellé se rapproche le plus de celui du produit à coder. Il classe ces produits par ordre de similarité probable à partir de l'analyse lexicale du libellé, et les propose dans cet ordre au codeur, en tenant compte d'un certain nombre de paramètres de priorité qui favorisent les produits que nous appelons «prototypes».

CREDOC/CODIT		CODAGE - MENU PRINCIPAL		CD0000
LIBELLE MANUEL	Codage d'un produit à partir d'un libelle manuel			
LIBELLES EXISTANTS	Codage des produits existant dans la base			
VALIDER THESAURUS	Validation des nouveaux mots du thesaurus			
CODAGE RECETTES	Codage complet des recettes et des ingrédients			
ANNUL				sion
OPTIO	Entrez un libelle complet de produit :			
YAOURT LAIT ENTIER NATURE POT EN VERRE				
Ok(PF4) Ecran Precedent(PF3)				

Alors deux possibilités se présentent.

A - L'un de ces produits correspond exactement au produit à coder. Par exemple, on pourra prendre ici comme modèle :

«Chambourcy pot verre/non sucré yaourts lait entier nature»

CREDOC/CODIT		AFFICHAGE DES PRODUITS SIMILAIRES		CD02 10
Produit à coder : 7000291 YAOURTS LAIT ENTIER NATURE POT EN VERRE				
Libelle	Comptref	Proto		
YAOURTS LAIT ENTIER NATURE	1199	1199		
CHAMBOURCY POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE	1200	0		
CHAMBOURCY POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE	1201	0		
CHAMBOURCY POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE	1202	0		
CHAMBOURCY POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE	1203	0		
DANONE POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 4	1298	0		
DANONE POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 8	1299	0		
DOUX PAYS POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE P	1745	0		
EUROMARCHE POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE	936	0		
GUIPE POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 3 X	975	0		
GUIPE POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 4 X	976	0		
LACTEL POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 3	1718	0		
LACTEL POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 6	1719	0		
MARQUES DISTRIBUTEURS-NON IDENTIFIEES POT VERRE SUCRE YAOURT	1083	0		
MARQUES DISTRIBUTEURS-NON IDENTIFIEES POT VERRE SUCRE YAOURT	1084	1		
Consulte descripteurs(PF4) Recherche libelle(F2) >				

Il suffit alors de valider ce produit comme «modèle identique». Le produit à coder hérite alors d'un seul coup de tous les codes du modèle. Le produit à coder est ajouté à OCABase. Il ne différera du modèle que par son libellé, son codeur, son origine, et les paramètres de gestion interne. Le codage a pris quelques

secondes, et il est parfaitement fidèle (deux opérateurs différents coderont la même chose).

Ce cas se présente heureusement d'autant plus souvent que la base s'agrandit.

B - Il n'existe pas de produit exactement identique au produit à coder. Le codeur examine, par un simple clic sur le nom du produit, chacun des produits déjà codés qui lui paraissent le plus proche. Les codes de ce produit apparaissent alors, en clair, sous la forme du fameux sandwich à facettes.

Produit à coder : 7000291

YAOURTS LAIT ENTIER NATURE POT EN VERRE

Produit codé : 1199 prototype

YAOURTS LAIT ENTIER NATURE

LANGUAL

INSEE

IFLS

Libellé	Code
PRODUIT A BASE DE LAIT FERMENTE	A0101
VACHE	B1201
LAIT	C0235
SEMI-SOLIDE A CONSISTANCE LISSE	E0119
-	E0000
TRANSFORMATION THERMIQUE PARTIELLE	F0018
PAS DE METHODE DE CUISSON APPLICABLE	G0003
-	G0000
FERMENTATION LACTIQUE	H0101
NON SUCRE	H0170
LAIT AJOUTE	H0184
-	H0000
REFROIDI	J0131
SANS MILIEU DE CONDITIONNEMENT	K0003
RECIPIENT OU EMBALLAGE NON SPECIFIE	M0001
SURFACE INCONNUE EN CONTACT AVEC L'ALIMENT	N0001
ALIMENT SANS PRECISION D'AGE	P0024
-	RGE00
-	S0000
-	T0000
-	Z0000
-	Z0000
-	Z0000
YAOURTS NATURE - SUCRES OU NON - Y COMPRIS YAOURTS A BOIRE	17411
YAOURT PRESURE ULTRA FRAIS	X9209

Puisque le modèle est proche, la plupart des facettes sont déjà bien codées. Il ne reste qu'à modifier à la marge certaines facettes.

Pour cela, le codeur n'a pas besoin de connaître les codes : il clique sur la facette à changer et le logiciel lui permet de se déplacer dans l'arborescence de la nomenclature correspondante jusqu'à ce qu'il ait trouvé le code qui lui convient.

Par exemple, ici, pour le type de conditionnement.

CREDOC/CODIT	CONSULTATION DEScriptEURS D'UN PRODUIT	CD0211
Produit à coder : 7000291 YAOURTS LAIT ENTIER NATURE POT EN VERRE		
Libelle	Code Langual	
REFROIDI	J0131	
SANS MILIEU DE CONDITIONNEMENT	K0003	
RECIPIENT OU EMBALLAGE NON SPECIFIE	M0001	
SURFACE INCONNUE EN CONTACT AVEC L'ALIMENT	N0001	
ALIMENT SANS PRECISION D'AGE	P0024	
-	P0000	
-	RGE00	
-	RGE00	
-	RGE00	
-	S0000	
-	T0000	
-	Z0000	
-	Z0000	
-	Z0000	
YAOURTS NATURE - SUCRES OU NON - Y COMPRIS YAOURTS A BOIRE	17411	
Accepter(PF4) Hierarchie(F2) Edition manuelle(F3) >		

Le codeur sélectionne dans la facette M le code du prototype «Récipient ou emballage non spécifié».

Le programme propose alors la liste des codes de même niveau. Puisque le codeur sait qu'il s'agit d'un pot en verre, il sélectionne «Récipient ou emballage de matière précise».

CREDOC/CODAGE	HIERARCHIE		CD0220
Libelle de la facette courante	Code	Niv	
RECIPIENT OU EMBALLAGE	M0100	1	
Libelles des facettes du niveau inferieur	Code	Niv	
RECIPIENT OU EMBALLAGE NON SPECIFIE	M0001	2	
SANS RECIPIENT OU EMBALLAGE	M0003	2	
RECIPIENT OU EMBALLAGE DIVERS	M0004	2	
RECIPIENT OU EMBALLAGE DE FORME PRECISE	M0195	2	
RECIPIENT OU EMBALLAGE DE MATIERE PRECISE	M0202	2	
Choisissez un descripteur avec les fleches.			
Niv. Sup.(PF4) Niv. Inf.(F2) Selection(F3) Aide(PF2) >			

Le logiciel lui propose alors différentes modalités, entre lesquelles il choisit «Récipient en verre».

CREDOC/CODAGE		HIERARCHIE		CD0220	
Libelle de la facette courante	Code	Niv			
RECIPIENT OU EMBALLAGE DE MATIERE PRECISE	M0202	2			
Libelles des facettes du niveau inférieur	Code	Niv			
RECIPIENT EN CERAMIQUE OU POTERIE	M0116	3			
RECIPIENT EN VERRE	M0180	3			
RECIPIENT METALLIQUE	M0151	3			
RECIPIENT EN TEXTILE	M0157	3			
RECIPIENT EN CARTON OU PAPIER	M0159	3			
RECIPIENT PLASTIQUE	M0172	3			
RECIPIENT EN BOIS	M0174	3			
RECIPIENT EN LAMINE	M0181	3			
RECIPIENT EN CIRE	M0209	3			

Choisissez un descripteur avec les flèches.

Niv. Sup.(PF4) Niv. Inf.(F2) Selection(F3) Aide(PF2) >

Il ne lui reste alors qu'à valider le code pour qu'il soit attribué au produit à coder.

CREDOC/CODIT		CONSULTATION DES DESCRIPTEURS D'UN PRODUIT		CD0211	
Produit à coder : 7000291					
YAOURTS LAIT ENTIER NATURE POT EN VERRE					
Libelle	Code	Langual			
REFROIDI	J0131				
SANS MILIEU DE CONDITIONNEMENT	K0003				
RECIPIENT EN VERRE	M0180				
SURFACE INCONNUE EN CONTACT AVEC L'ALIMENT	N0001				
ALIMENT SANS PRECISION D'AGE	P0024				
-	P0000				
-	RGE00				
-	RGE00				
-	RGE00				
-	S0000				
-	T0000				
-	Z0000				
-	Z0000				
-	Z0000				
YAOURTS NATURE - SUCRES OU NON - Y COMPRIS YAOURTS A BOIRE	17411				

Accepter(PF4) Hierarchie(F2) Edition manuelle(F3) >

Pour changer ou construire une recette, le codeur n'a qu'à indiquer en clair le nom de chaque ingrédient. Le logiciel va chercher, comme pour le codage du produit entier, l'ingrédient le plus proche déjà codé dans la base. Le codeur choisit parmi ceux qui lui sont proposés. Il peut consulter par simple clic le codage de ces ingrédients pour s'assurer de la pertinence de son choix.

Par ailleurs, le logiciel est fait pour pouvoir coder «en série». Dans le cas où on veut coder de nombreux produits à la suite, plutôt que de taper chaque nom de produit à coder, on donne à la machine un fichier contenant la liste des libellés à coder. Le logiciel les code alors les uns à la suite des autres. Mais, lorsqu'un produit est codé, il examine la liste de tous les produits restant à coder pour voir s'il ne serait pas possible de coder d'un bloc toute une liste de produits identiques. C'est très souvent le cas lorsque l'on code des références de produits manufacturés.

Par exemple, sur le marché des yaourts, on aura 33 références de yaourts nature au lait entier qui ne diffèrent que par la marque et le «par combien» (par 4, par 6, par 12...), mais qui sont physiquement identiques (même procédé de fabrication, même conditionnement individuel en pot en verre avec operculage en aluminium).

Le logiciel présente alors au codeur la liste des références qui lui semblent analogues, et propose au codeur d'en coder tout ou partie de la même façon que le modèle initial.

871	AUTRES MARQUES-NON IDENTIFIEES POT VERRE / NON SUCRE YAOURTS LAIT
872	AUTRES MARQUES-NON IDENTIFIEES POT VERRE / NON SUCRE YAOURTS LAIT
873	AUTRES MARQUES-NON IDENTIFIEES POT VERRE / NON SUCRE YAOURTS LAIT
874	AUTRES MARQUES-NON IDENTIFIEES POT VERRE / NON SUCRE YAOURTS LAIT
936	EUROMARCHE POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 3
942	CERCLE ROUGE POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 3
975	GUIPE POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 3
976	GUIPE POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 4
992	MONT JOLY POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 3
1199	CHAMBOURCY POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 1
1200	CHAMBOURCY POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 3
1201	CHAMBOURCY POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 4
1202	CHAMBOURCY POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 6
1203	CHAMBOURCY POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 8
1298	DANONE POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 4
1299	DANONE POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 8
1336	STENVAL POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE FAMILIAL
1429	NOVA POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 2 X 2
1430	NOVA POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 3
1431	NOVA POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 4
1432	NOVA POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 6
1565	YOPLAIT POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 2
1566	YOPLAIT POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 3
1567	YOPLAIT POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 4
1568	YOPLAIT POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 6
1569	YOPLAIT POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 8
1609	SENOBLE POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 3
1610	SENOBLE POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE
1718	LACTEL POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 3 X 3
1719	LACTEL POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE X 6 X 6
1745	DOUX PAYS POT VERRE / NON SUCRE YAOURTS LAIT ENTIER NATURE POT IN
1768	RIANS POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 2
1769	RIANS POT VERRE NON SUCRE YAOURTS LAIT ENTIER NATURE X 3

Ceci permet d'accélérer encore la vitesse du codage.

Il existe d'autres subtilités du logiciel, mais il serait trop long de les présenter ici et elles ne présentent pas d'intérêt théorique important.

4. PROJETS

Au niveau français, des travaux concernant la partie linguistique de la codification sont prévus dans le cadre des travaux de l'OCA. Car, si une partie difficile du chemin a été faite en résolvant les problèmes théoriques liés à la modélisation opératoire des aliments, beaucoup reste à faire sur la base de données pour que l'OCA puisse remplir pleinement le rôle qui lui a été confié. En particulier on envisage d'avoir recours à une meilleure prise en compte du sens en utilisant des graphes conceptuels, la reconnaissance des noms composés, et la compréhension des grammaires alimentaires. Des recherches sont en cours.

Ce type de développement est nécessaire à l'OCA pour plusieurs raisons. D'abord, parce qu'ils permettront de mieux gérer l'intégrité de la base, de contrôler la qualité de l'information, et de l'extraire plus facilement pour répondre aux demandes des utilisateurs. Ensuite, la mise au point de logiciels de codification puissants et conviviaux est une condition indispensable à l'amélioration de la précision et l'harmonisation progressive des données d'enquêtes en France et en Europe.

BIBLIOGRAPHIE

BENGUIGUI, G. «Les besoins des objets de consommation et les groupes sociaux», *Epistémologie sociologique*, n° 15-16, 1973

BORGES, J.L., «La langue analytique de John Wilkins». *Enquêtes*, 1937-1942, NRF, Gallimard. 1957.

Center for Food Safety and Applied Nutrition, *Languag Vocabulary User's Manual*, Rev. December 1990.

Center for Food Safety and Applied Nutrition, *Languag an automated method for describing, capturing an retrieving data about food*, Codata conference, Columbus, Ohio, USA, 1990.

FEINBERG, M., IRELAND-RIPERT, J. FAVIER, J.-C. «Languag : un langage international pour la description structurée des aliments», *Sciences des Aliments*, 11 (1991) 189-210

HEISENBERG, W. La nature dans la physique contemporaine. Gallimard, 1962.

LAHLOU, S. Le produit nouveau : un concept flou. *Consommation* N° 2, 1985-86.

LAHLOU, S. Les comportements alimentaires des Français. Rapport Crédoc, 1988.

LAHLOU, S. La système-compatibilité. *Cahiers de recherche du Crédoc*, n° 4, 1990.

LAHLOU, S., MAFFRE, J., BEAUDOUIN, V. La codification des objets complexes : réflexions théoriques

et application à un corpus de 8000 produits alimentaires. *Cahier de recherche du Crédoc*, n° 23, décembre 1991.

LAHLOU S., BEAUDOUIN V., CALAMASSI-TRAN G., EVANS C., GILLET C., LION S., MAFFRE J., VERHEYDEN G. Rapport pour l'observatoire des consommations alimentaires. Etat d'avancement des travaux de la base de données mise en place par le Crédoc à la fin de la deuxième phase : décembre 1991. Rapport Crédoc 52 p. + 13 annexes.

VON LINNE, C. *Voyage en Laponie*, Editions de la différence. Paris, 1983.

MONOD-HERTZEN, G. *L'analyse dimensionnelle et l'épistémologie*. Coll. recherches interdisciplinaires. Editions Maloine-Doine, Paris, 1976.

MOREL, B. «Réflexions philosophiques à partir de la «méthodologie ouverte»». in J. Parain-vial, ed. *Les difficultés de la quantification et de la mesure. Actes du colloque de l'université de Dijon «Méthodologie comparée des sciences»*. Maloine, Paris, 1981. pp. 251-263.

POIRIER, R., «Nature et réalité de l'objet». in J. Parain-Vial, ed. *Les difficultés de la quantification et de la mesure. Actes du colloque de l'université de Dijon «Méthodologie comparée des sciences»*. Maloine, Paris, 1981. pp. 275-293.

QUINE, W.V.O. *Le mot et la chose*. Flammarion, Paris, 1977.

ROSCH, E.R. Universals and Specifics in Human Categorization, in R. Brislin, S. Bosner & W. Lonner eds. *Cross Cultural Perspectives on Learning*, Hasted, New-York, 1975.

ROSCH, E.R., MERVIS, C.B., GRAY, W., JOHNSON, D.N., BOYES-BRAEM, P. Basic Objects in Natural Categories, *Cognitive Psychology*, 1976, 8, pp. 382-489.

SAGET, H. Nature et limites de la quantification. in J. Parain-Vial, ed. *Les difficultés de la quantification et de*

la mesure. Actes du colloque de l'université de Dijon «Méthodologie comparée des sciences». Maloine, Paris, 1981. pp. 265-274.

SERRES, M. *Le système de Leibniz et ses modèles mathématiques*, PUF, Paris, 1968.

WITTGENSTEIN, L. *Tractatus Logico-philosophicus*, Gallimard, Paris, 1961.

WITTGESTEIN, L. *Le brun et le cahier bleu*, Trad fr. Gallimard, 1965.